

06. 8. 2004

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

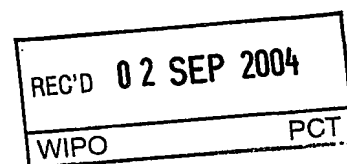
別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日  
Date of Application: 2003年 8月20日

出 願 番 号  
Application Number: 特願2003-295758  
[ST. 10/C]: [JP2003-295758]

出 願 人  
Applicant(s): 日本電気株式会社

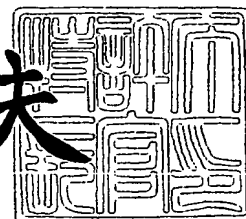


**PRIORITY  
DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1 (a) OR (b)

2004年 6月14日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



【書類名】 特許願  
【整理番号】 33509999  
【提出日】 平成15年 8月20日  
【あて先】 特許庁長官殿  
【国際特許分類】 H04L 12/00  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 下西 英之  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 村瀬 勉  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 山崎 康広  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 長谷川 洋平  
【特許出願人】  
    【識別番号】 000004237  
    【氏名又は名称】 日本電気株式会社  
【代理人】  
    【識別番号】 100088812  
    【弁理士】  
    【氏名又は名称】 ▲柳▼川 信  
【手数料の表示】  
    【予納台帳番号】 030982  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1  
    【包括委任状番号】 9001833

**【書類名】 特許請求の範囲****【請求項1】**

複数のレイヤにおける輻輳制御処理とパケット出力制御処理とを含むセッション中継処理を行うセッション中継装置であって、前記複数のレイヤ各々が前記輻輳制御情報の作成のみを行い、前記パケット出力制御処理をIP (Internet Protocol) レイヤのスケジューラに集約することを特徴とするセッション中継装置。

**【請求項2】**

前記複数のレイヤに対応する受信バッファ及び送信バッファを前記IPレイヤに対応する送信バッファに集約することを特徴とする請求項1記載のセッション中継装置。

**【請求項3】**

送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記受信端末と前記送信端末との間の通信を実現するセッション中継装置であって、

前記送信端末に向けたセッションからのデータを受信する受信セッション処理手段と、前記受信端末に向けたセッションへとデータを送信する送信セッション処理手段と、前記送信端末へと出力するデータを一時蓄えておく送信バッファと、前記送信バッファからのパケット出力を制御するパケットスケジューラと、前記パケットスケジューラの制御に 응답して前記送信バッファに蓄えられたデータを出力制御する出力制御手段とを有し、

前記送信セッション処理手段において当該レイヤで出力が許可されているデータ量を計算し、これに基づいて前記パケットスケジューラが前記パケット出力を制御することを特徴とするセッション中継装置。

**【請求項4】**

前記受信セッション処理手段は、TCP (Transmission Control Protocol) セッションからのデータ受信処理を行い、

前記送信セッション処理手段は、前記TCPセッションへのデータ出力処理を行い、TCP ウインドフロー制御によって決定される出力可能なデータ量を前記パケットスケジューラに通知し、

前記パケットスケジューラは、通知されたデータ量を基にスケジューリング処理を行うことを特徴とする請求項3記載のセッション中継装置。

**【請求項5】**

前記パケットスケジューラは、前記セッションに割り当てられた帯域及び帯域比率を少なくとも含む通信資源割り当て方針と、前記送信セッション処理手段から通知された送信可能データ量と、前記送信バッファ内に蓄えられているデータ量とを基にパケット出力を行うセッションを決定し、前記セッション各々からのデータ出力を制御することを特徴とする請求項3または請求項4記載のセッション中継装置。

**【請求項6】**

前記セッション各々における未使用の通信資源を蓄えておく蓄積手段を前記パケットスケジューラに含み、

前記パケットスケジューラは、前記通信資源が必要になった際に前記蓄積手段に蓄えておいた通信資源を用いて通信を行うことを特徴とする請求項3から請求項5のいずれか記載のセッション中継装置。

**【請求項7】**

前記パケットスケジューラは、前記送信バッファ内に出力すべきデータがあり、前記出力制御手段からの出力許可データ量の制限によって未使用となった前記通信資源の帯域のみを前記蓄積手段に蓄えておくことを特徴とする請求項6記載のセッション中継装置。

**【請求項8】**

前記送信セッションの制御パラメータを動的に変更する手段を含み、前記パケットスケジューラからのデータ出力状況に応じて前記制御パラメータを変更することを特徴とする請求項3から請求項7のいずれか記載のセッション中継装置。

**【請求項9】**

前記セッションの未使用帯域が大きくなった時に当該セッションの制御パラメータを当該セッションからの出力帯域が小さくなる方向に変更し、前記セッションの未使用帯域が小さくなった時に当該セッションの制御パラメータを当該セッションからの出力帯域が大きくなる方向に変更し、前記制御パラメータの変更によって輻輳が発生した時に前記制御パラメータの変更を停止することを特徴とする請求項 8 記載のセッション中継装置。

【請求項 10】

前記セッション各々に割り当てた帯域及び帯域比率を少なくとも含む通信資源割り当て量を動的に変更する手段を含み、

前記パケットスケジューラからのデータ出力状況及び前記出力制御手段から通知される送信可能データ量に応じて前記制御パラメータを変更することを特徴とする請求項 9 記載のセッション中継装置。

【請求項 11】

前記セッションの未使用帯域が大きくなった時に当該セッションの割り当て資源を減少させ、前記セッションの未使用帯域が小さくなった時に当該セッションの割り当て資源をその初期値を上限として増加させるとともに、前記出力制御手段から通知される送信可能データ量及びその平均のいずれかによって前記割り当て資源を増減することを特徴とする請求項 10 記載のセッション中継装置。

【請求項 12】

前記送信端末からのセッションの送信処理を制御する帯域、送信可否、送信可能データ量を少なくとも含む送信制御情報を制御する受信レート制御手段を含み、前記送信バッファの空き容量及び前記パケットスケジューラからの情報に応じて前記送信端末への送信制御情報の変更及び生成のいずれかを行うことを特徴とする請求項 3 から請求項 10 のいずれか記載のセッション中継装置。

【請求項 13】

前記パケットスケジューラからのパケット出力情報を受信する手段と、パケット出力によって変更された前記送信バッファの空き容量を調べる手段とを含み、パケット出力後に前記送信バッファの空き容量が一定量以上となった時に前記送信端末に対して送達確認パケットを送信して送信再開を促すことを特徴とする請求項 12 記載のセッション中継装置。

【請求項 14】

前記送信バッファの空き容量及びその平均の少なくとも一方を調べる手段を含み、前記空き容量に応じて前記送信端末に対して送信帯域の減少を指示することを特徴とする請求項 12 記載のセッション中継装置。

【請求項 15】

送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記送信端末と前記受信端末との間の通信を実現するセッション中継装置であって、

複数のレイヤに対応して設けられかつ前記送信端末に向けたセッションからのデータを受信する受信セッション処理手段と、前記複数のレイヤに対応して設けられかつ前記受信端末に向けたセッションへとデータを送信する送信セッション処理手段と、前記送信端末へと出力するデータを一時蓄えておく送信バッファと、前記送信バッファからのパケット出力を制御するパケットスケジューラとを有し、

前記送信セッション処理手段各々において当該レイヤで出力が許可されているデータ量を計算し、前記複数のレイヤ全てで共通に許可されるデータ量に基づいて前記パケットスケジューラが前記パケット出力を制御することを特徴とするセッション中継装置。

【請求項 16】

前記レイヤとして輻輳制御を行うレイヤのひとつとして iSCSI (internet Small Computer System Interface) レイヤを含み、当該 iSCSI レイヤにおいて前記受信端末から受信する受信可能データ量を基に送信可能データ量を決定することを特徴とする請求項 15 記載のセッション中継装置。

**【請求項 17】**

前記パケットスケジューラからのパケット出力情報を受信する手段と、パケット出力によって変更された前記送信バッファの空き容量を調べる手段とを含み、パケット出力後に前記送信バッファの空き容量が一定量以上となった時に前記送信端末に対して前記受信可能データ量を生成して送信再開を促すことを特徴とする請求項 15 または請求項 16 記載のセッション中継装置。

**【請求項 18】**

前記受信セッション処理手段は、受信したパケットを前記送信バッファに直接格納し、前記送信バッファから直接出力することを特徴とする請求項 3 から請求項 17 のいずれか記載のセッション中継装置。

**【請求項 19】**

アプリケーションプログラムから前記送信バッファへとデータの書き込みを行い、受信したデータを前記アプリケーションプログラムへと渡すことを特徴とする請求項 3 から請求項 18 のいずれか記載のセッション中継装置。

**【請求項 20】**

複数のレイヤにおける輻輳制御処理とパケット出力制御処理とを含むセッション中継処理を行うセッション中継装置のセッション中継方法であって、前記複数のレイヤ各々で前記輻輳制御情報の作成のみを行い、前記パケット出力制御処理を IP (Internet Protocol) レイヤのスケジューラに集約することを特徴とするセッション中継方法。

**【請求項 21】**

前記複数のレイヤに対応する受信バッファ及び送信バッファを前記 IP レイヤに対応する送信バッファに集約することを特徴とする請求項 20 記載のセッション中継方法。

**【請求項 22】**

送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記受信端末と前記送信端末との間の通信を実現するセッション中継装置のセッション中継方法であって、前記セッション中継装置側に、前記送信端末に向けたセッションからのデータを受信する受信セッション処理と、前記受信端末に向けたセッションへとデータを送信する送信セッション処理と、前記送信端末へと出力するデータを送信バッファに一時蓄えておく処理と、前記送信バッファからのパケット出力をパケットスケジューラにて制御する処理と、前記パケットスケジューラの制御に応答して前記送信バッファに蓄えられたデータを出力制御手段にて出力制御する処理とを有し、前記送信セッション処理において当該レイヤで出力が許可されているデータ量を計算し、これに基づいて前記パケットスケジューラが前記パケット出力を制御することを特徴とするセッション中継方法。

**【請求項 23】**

前記受信セッション処理は、TCP (Transmission Control Protocol) セッションからのデータ受信処理を行い、

前記送信セッション処理は、前記 TCP セッションへのデータ出力処理を行い、TCP ウインドフロー制御によって決定される出力可能なデータ量を前記パケットスケジューラに通知することで、前記パケットスケジューラが通知されたデータ量を基にスケジューリング処理を行うことを特徴とする請求項 22 記載のセッション中継方法。

**【請求項 24】**

前記パケットスケジューラは、前記セッションに割り当てられた帯域及び帯域比率を少なくとも含む通信資源割り当て方針と、前記送信セッション処理手段から通知された送信可能データ量と、前記送信バッファ内に蓄えられているデータ量とを基にパケット出力を行うセッションを決定し、前記セッション各々からのデータ出力を制御することを特徴とする請求項 22 または請求項 23 記載のセッション中継方法。

**【請求項 25】**

前記セッション各々における未使用の通信資源を蓄えておく蓄積手段を前記パケットス

ケジューラに含み、前記パケットスケジューラにおいて前記通信資源が必要になった際に前記蓄積手段に蓄えておいた通信資源を用いて通信を行うことを特徴とする請求項 22 から請求項 24 のいずれか記載のセッション中継方法。

【請求項 26】

前記パケットスケジューラは、前記送信バッファ内に出力すべきデータがあり、前記出力制御手段からの出力許可データ量の制限によって未使用となった前記通信資源の帯域のみを前記蓄積手段に蓄えておくことを特徴とする請求項 25 記載のセッション中継方法。

【請求項 27】

前記送信セッションの制御パラメータを動的に変更する手段が、前記パケットスケジューラからのデータ出力状況に応じて前記制御パラメータを変更することを特徴とする請求項 22 から請求項 26 のいずれか記載のセッション中継方法。

【請求項 28】

前記セッションの未使用帯域が大きくなった時に当該セッションの制御パラメータを当該セッションからの出力帯域が小さくなる方向に変更し、前記セッションの未使用帯域が小さくなった時に当該セッションの制御パラメータを当該セッションからの出力帯域が大きくなる方向に変更し、前記制御パラメータの変更によって輻輳が発生した時に前記制御パラメータの変更を停止することを特徴とする請求項 27 記載のセッション中継方法。

【請求項 29】

前記セッション各々に割り当てた帯域及び帯域比率を少なくとも含む通信資源割り当て量を動的に変更する手段が、前記パケットスケジューラからのデータ出力状況及び前記出力制御手段から通知される送信可能データ量に応じて前記制御パラメータを変更することを特徴とする請求項 28 記載のセッション中継方法。

【請求項 30】

前記セッションの未使用帯域が大きくなった時に当該セッションの割り当て資源を減少させ、前記セッションの未使用帯域が小さくなった時に当該セッションの割り当て資源をその初期値を上限として増加させるとともに、前記出力制御手段から通知される送信可能データ量及びその平均のいずれかによって前記割り当て資源を増減することを特徴とする請求項 29 記載のセッション中継方法。

【請求項 31】

前記送信端末からのセッションの送信処理を制御する帯域、送信可否、送信可能データ量を少なくとも含む送信制御情報を制御する受信レート制御手段が、前記送信バッファの空き容量及び前記パケットスケジューラからの情報に応じて前記送信端末への送信制御情報の変更及び生成のいずれかを行うことを特徴とする請求項 22 から請求項 29 のいずれか記載のセッション中継方法。

【請求項 32】

パケット出力後に前記送信バッファの空き容量が一定量以上となった時に前記送信端末に対して送達確認パケットを送信して送信再開を促すことを特徴とする請求項 31 記載のセッション中継方法。

【請求項 33】

前記送信バッファの空き容量及びその平均の少なくとも一方を調べる手段で調べられた前記空き容量に応じて前記送信端末に対して送信帯域の減少を指示することを特徴とする請求項 31 記載のセッション中継方法。

【請求項 34】

送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記送信端末と前記受信端末との間の通信を実現するセッション中継装置のセッション中継方法であって、前記セッション中継装置側に、複数のレイヤ各々において前記送信端末に向けたセッションからのデータを受信する受信セッション処理と、前記複数のレイヤ各々において前記受信端末に向けたセッションへとデータを送信する送信セッション処理と、前記送信端末へと出力するデータを送信バッファに一時蓄えておく処理と、前記送信バッファからのパケット出力をパケットスケジューラにて制御する処理とを有し、

前記送信セッション処理各々において当該レイヤで出力が許可されているデータ量を計算し、前記複数のレイヤ全てで共通に許可されるデータ量に基づいて前記パケットスケジューラが前記パケット出力を制御することを特徴とするセッション中継方法。

【請求項 35】

前記レイヤとして輻輳制御を行うレイヤのひとつとして iSCSI (internet Small Computer System Interface) レイヤを含み、当該 iSCSI レイヤにおいて前記受信端末から受信する受信可能データ量を基に送信可能データ量を決定することを特徴とする請求項 34 記載のセッション中継方法。

【請求項 36】

パケット出力後に前記送信バッファの空き容量が一定量以上となった時に前記送信端末に対して前記受信可能データ量を生成して送信再開を促すことを特徴とする請求項 34 または請求項 35 記載のセッション中継方法。

【請求項 37】

前記受信セッション処理は、受信したパケットを前記送信バッファに直接格納し、前記送信バッファから直接出力することを特徴とする請求項 22 から請求項 36 のいずれか記載のセッション中継方法。

【請求項 38】

アプリケーションプログラムから前記送信バッファへとデータの書き込みを行い、受信したデータを前記アプリケーションプログラムへと渡すことを特徴とする請求項 22 から請求項 37 のいずれか記載のセッション中継方法。

【書類名】 明細書

【発明の名称】 セッション中継装置及びそれに用いるセッション中継方法

【技術分野】

【0001】

本発明はセッション中継装置及びそれに用いるセッション中継方法に関し、特にTCP (Transmission Control Protocol) セッション間でデータの中継を行う装置に関する。

【背景技術】

【0002】

一般的に、通信アプリケーションにおいては、送信端末と受信端末との間で通信セッションを確立し、確立したセッション上で通信を行う。しかしながら、送信端末と受信端末との間の伝播遅延時間が非常に長い場合、もしくは有線と無線のように特性の異なるネットワークを横断して通信する場合には、送信端末と受信端末との間の通信のスループットが低下する。

【0003】

この問題を解決するための方法としては、送信端末と受信端末との間を一つのセッションで通信を行うのではなく、送信端末と受信端末との間に中継装置を設置し、送信端末から中継装置へのセッションと中継装置から受信端末へのセッションの2つのセッション間でデータの中継を行うことによって通信を行う方式が存在する（例えば、特許文献1、2、非特許文献1参照）。

【0004】

中継装置においては、伝播遅延時間が非常に長い場合でもスループットを低下させないために、大きな送信バッファを持つことが有効であることが知られている。しかしながら、大きな送信バッファを持った場合、特にTCPのスロースタート (Slow Start) 時において、パケットがバースト的に出力されるためにネットワークの輻輳を発生させ、結果としてスループットの低下を招く場合がある。

【0005】

このような問題を解決する方法としては、TCPセッションからバースト的にパケットを出力しないように出力するパケットの間隔を調整する方法がある（例えば、非特許文献2参照）。

【0006】

また、他の方法としては、TCPセッションからの出力パケットをキューイングしてスケジューラで出力制御を行うことによって、任意の帯域でのパケット出力が可能となり、他のTCPセッションのスループットを制限し、特定のセッションのスループットを向上させることも可能である。

【0007】

セッションの中継を行う場合、最も問題となることは、一般的にTCPセッションの処理において処理負荷が高く、高速な中継処理が難しいという問題がある。また、TCPセッションからのパケット出力間隔を調整する場合には、処理負荷がさらに高くなる。

【0008】

上記のTCPセッションの処理を高速化する方式としては、以下に示すような方式が存在する。第1の方式としては、OS (Operation System) のカーネルとアプリケーションプログラムとの間のデータ転送にゼロコピー方式を用いることによって、中継装置の処理負荷を下げる方式がある。この第1の方式では、実際にデータの送受信処理を行うカーネルプログラムと、データの中継処理を行うアプリケーションプログラムとの間で、データの受け渡しを行う際、物理的なデータのコピーを行わず、ページマッピングによる仮想的なデータの移動を行う。

【0009】

第2の方式としては、中継処理を行う際、再送制御のみを行う方式がある。この第2の方式では、中継装置でセッションの終端を行わず、送受信端末間で1本のセッションを設



定する。そして、中継装置においては送信端末から受信したパケットを受信端末へと出力する際、単にパケットを転送するだけでなく、パケットを中継装置内に保存しておく。

#### 【0010】

中継装置では受信端末から送信端末に返されるACK (acknowledgement) パケットを監視し、もし中継装置と受信端末との間でパケット廃棄を検出すると、保存しておいたパケットを受信端末に出力することで、送信端末からのパケットの再送を防ぎ、パケット廃棄による送信端末と受信端末との間のスループットの低下を防いでいる。この第2の方式では、セッション間の中継処理を行わず、再送処理を行うのみであり、中継処理を行う場合に比べて中継装置の負荷を下げるができる。

#### 【0011】

第3の方式としては、中継の必要が無いセッションの中継処理を行わず、中継を行う必要があるセッションのみ中継処理を行うことで、中継装置の処理負荷を下げる方式がある。例えば、この方式としては、セッションのスループットを計測し、中継処理を行うことで、スループットが向上するセッションのみを選択して中継処理を行う方式がある（例えば、特許文献3参照）。

#### 【0012】

また、上記の方式としては、HTTP (Hyper Text Transfer Protocol) のリクエストを送信する際に中継処理を行わず、データ転送を行う際のみ中継処理を行う方式がある（例えば、特許文献4参照）。

#### 【0013】

【特許文献1】特開平11-252179号公報

【特許文献2】特開2002-281104号公報

【特許文献3】特開平11-112576号公報

【特許文献4】特開2002-312261号公報

【非特許文献1】jay Bakre and B. R. Badrinath, "I-TCP; Indirect TCP for Mobile Host", Department of Computer Science Rutgers University, DSC-TR-314, 1994 (<http://www.it.iitb.ac.in/it644/papers/i-tcp.pdf>)

【非特許文献2】A. Agarwal, S. Savage, and T. Anderson. "Understanding the Performance of TCP Pacing", in proceedings of IEEE INFOCOM'2000

#### 【発明の開示】

#### 【発明が解決しようとする課題】

#### 【0014】

上述した従来のセッション中継方式では、第1の方式の場合、データコピー以外の処理負荷も高いという問題がある。TCP pacing方式やスケジューラによる出力制御を行う場合、TCPの処理負荷に加えて、パケット出力制御の負荷が加わるため、全体の処理負荷が高くなる。

#### 【0015】

また、TCPセッションの中継だけではなく、例えばiSCSI (internet Small Computer System Interface) のような上位レイヤプロトコルの中継処理を行う場合、さらに上位レイヤでの輻輳制御に基づくパケット出力制御が加わり、さらに処理負荷が高くなる。特に、中継するセッション数が多い場合には、セッション間でパケットの出力制御を行う処理の負荷が高くなる。

#### 【0016】

さらに、従来のセッション中継方式では、第1の方式の場合、パケット長に制限があるという問題がある。TCPレイヤとアプリケーションとの間でページマッピングによる仮想的なデータの移動を行うためには、CPU (中央処理装置) のページサイズとパケット

長とが一致している必要がある。

【0017】

一般的に、セッション中継装置は送受信端末とは独立して存在し、送信端末及び受信端末が用いるパケット長を予め想定しておくことはできないため、帯域制御装置のページサイズと異なるパケット長を用いる送信端末及び受信端末に対しては、ゼロコピーによる処理負荷低減が期待できない。

【0018】

一方、従来のセッション中継方式では、第2の方式の場合、パケット廃棄以外の要因によるスループット低下を改善することができないという問題がある。セッションの中継処理を行う場合には、セッション中継装置では送信端末からのパケット受信に対してすぐさま送信端末に対して受信確認のためのACKパケットを返送するが、この第2の方式では受信端末が送信端末に対してACKパケットを返すのみであり、セッション中継装置自身はACKパケットを返さない。そのため、伝播遅延が大きな環境では第2の方式によるスループット向上の効果が小さい。

【0019】

また、従来のセッション中継方式では、第3の方式の場合、中継する必要があるセッション数が多い場合に処理負荷が軽減されないという問題がある。

そこで、本発明の目的は上記の問題点を解消し、中継するセッション数が多く、パケットの出力制御を行う場合でも、高速にセッション間の中継処理を行うことができるセッション中継装置及びそれに用いるセッション中継方法を提供することにある。

【課題を解決するための手段】

【0020】

本発明によるセッション中継装置は、複数のレイヤにおける輻輳制御処理とパケット出力制御処理とを含むセッション中継処理を行うセッション中継装置であって、前記複数のレイヤ各々が前記輻輳制御情報の作成のみを行い、前記パケット出力制御処理をIP (Internet Protocol) レイヤのスケジューラに集約している。

【0021】

本発明による他のセッション中継装置は、送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記受信端末と前記受信端末との間の通信を実現するセッション中継装置であって、

前記送信端末に向けたセッションからのデータを受信する受信セッション処理手段と、前記受信端末に向けたセッションへとデータを送信する送信セッション処理手段と、前記送信端末へと出力するデータを一時蓄えておく送信バッファと、前記送信バッファからのパケット出力を制御するパケットスケジューラと、前記パケットスケジューラの制御にตอบสนองして前記送信バッファに蓄えられたデータを出力制御する出力制御手段とを備え、

前記送信セッション処理手段において当該レイヤで出力が許可されているデータ量を計算し、これに基づいて前記パケットスケジューラが前記パケット出力を制御している。

【0022】

本発明による別のセッション中継装置は、送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記送信端末と前記受信端末との間の通信を実現するセッション中継装置であって、

複数のレイヤに対応して設けられかつ前記送信端末に向けたセッションからのデータを受信する受信セッション処理手段と、前記複数のレイヤに対応して設けられかつ前記受信端末に向けたセッションへとデータを送信する送信セッション処理手段と、前記送信端末へと出力するデータを一時蓄えておく送信バッファと、前記送信バッファからのパケット出力を制御するパケットスケジューラとを備え、

前記送信セッション処理手段各々において当該レイヤで出力が許可されているデータ量を計算し、前記複数のレイヤ全てで共通に許可されるデータ量に基づいて前記パケットスケジューラが前記パケット出力を制御している。

【0023】

本発明によるセッション中継方法は、複数のレイヤにおける輻輳制御処理とパケット出力制御処理とを含むセッション中継処理を行うセッション中継装置のセッション中継方法であって、前記複数のレイヤ各々で前記輻輳制御情報の作成のみを行い、前記パケット出力制御処理をIP (Internet Protocol) レイヤのスケジューラに集約している。

#### 【0024】

本発明による他のセッション中継方法は、送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記受信端末と前記受信端末との間の通信を実現するセッション中継装置のセッション中継方法であって、前記セッション中継装置側に、前記送信端末に向けたセッションからのデータを受信する受信セッション処理と、前記受信端末に向けたセッションへとデータを送信する送信セッション処理と、前記送信端末へと出力するデータを送信バッファに一時蓄えておく処理と、前記送信バッファからのパケット出力をパケットスケジューラにて制御する処理と、前記パケットスケジューラの制御に応答して前記送信バッファに蓄えられたデータを出力制御手段にて出力制御する処理とを備え、前記送信セッション処理において当該レイヤで出力が許可されているデータ量を計算し、これに基づいて前記パケットスケジューラが前記パケット出力を制御している。

#### 【0025】

本発明による別のセッション中継方法は、送信端末に向けたセッションと受信端末に向けたセッションとの間でデータの中継を行うことで前記送信端末と前記受信端末との間の通信を実現するセッション中継装置のセッション中継方法であって、前記セッション中継装置側に、複数のレイヤ各々において前記送信端末に向けたセッションからのデータを受信する受信セッション処理と、前記複数のレイヤ各々において前記受信端末に向けたセッションへとデータを送信する送信セッション処理と、前記送信端末へと出力するデータを送信バッファに一時蓄えておく処理と、前記送信バッファからのパケット出力をパケットスケジューラにて制御する処理とを備え、前記送信セッション処理各々において当該レイヤで出力が許可されているデータ量を計算し、前記複数のレイヤ全てで共通に許可されるデータ量に基づいて前記パケットスケジューラが前記パケット出力を制御している。

#### 【0026】

すなわち、本発明のセッション中継装置は、TCP (Transmission Control Protocol) レイヤにおいてパケット出力の制御を行わず、TCPレイヤでパケット出力のための制御情報を生成するのみとし、IP (Internet Protocol) レイヤにおいてパケットスケジューラを用いてパケット出力の制御を行う。

#### 【0027】

また、本発明のセッション中継装置では、iSCSI (internet Small Computer System Interface) のような上位レイヤプロトコルにおける輻輳制御に関しても、制御情報の作成のみを上位レイヤで行い、実際のパケット出力制御にはIPレイヤのパケットスケジューラを用いる。

#### 【0028】

これによって、本発明のセッション中継装置では、複数のレイヤにおけるデータ出力制御を統合することができるため、パケット出力に関する処理負荷を軽減することが可能となる。

#### 【0029】

さらに、本発明のセッション中継装置では、アプリケーションへのデータコピーを行わないため、ページマッピングによるデータの移動の必要がない。すなわち、受信パケットはTCPレイヤやその他の上位レイヤ、アプリケーションの受信バッファに格納されることなく、またTCPレイヤやその他の上位レイヤ、アプリケーションの送信バッファに格納に格納されることもなく、直接IPレイヤの送信バッファに格納される。そのため、本発明のセッション中継装置では、セッション中継装置内のデータの移動は発生せず、ペー

ジマッピングによるデータの移動の必要がない。

【0030】

さらにまた、本発明のセッション中継装置では、中継処理をパケットの再送制御にのみを行うのではなく、セッションを一旦終端して完全な中継処理を行っているため、パケット廃棄以外の要因によるスループット低下を改善することが可能となる。

【0031】

本発明のセッション中継装置では、セッション間の中継処理を高速化することで、中継セッション数が多い場合でも高速に処理を行うことが可能となる。

【発明の効果】

【0032】

本発明は、以下に述べるような構成及び動作とすることで、セッション中継処理の高速化を実現することができるという効果が得られる。

【発明を実施するための最良の形態】

【0033】

次に、本発明の実施の形態について図面を参照して説明する。図1は本発明の実施の形態によるセッション中継装置の構成を示すブロック図である。図1において、本発明の実施の形態によるセッション中継装置1はセッション識別部11と、セッション中継部12-1～12-Nと、出力制御部14とから構成され、受信端末2及び送信端末3に接続されている。

【0034】

まず、送信端末3から受信端末2へデータを送る場合、送信端末3からのデータパケットはセッション中継部12-1の受信セッション処理部（図示せず）で処理され、その結果、ACK (acknowledgement) パケットが送信端末3へと返信される。

【0035】

セッション中継部12-1の受信セッション処理部で受取られたデータはセッション中継部12-1の送信セッション処理部（図示せず）へと送られ、ここから受信端末2へとデータパケットが送信される。これに対して、受信端末2が返信したACKパケットはセッション中継部12-1の送信セッション処理部で処理される。

【0036】

同様に、受信端末2から送信端末3へデータを送る場合、受信端末2からのデータパケットはセッション中継部12-2の受信セッション処理部（図示せず）で処理され、その結果、ACKパケットが受信端末2へと返信される。

【0037】

セッション中継部12-2の受信セッション処理部で受取られたデータはセッション中継部12-2の送信セッション処理部（図示せず）へと送られ、ここから送信端末3へとデータパケットが送信される。これに対して、送信端末3が返信したACKパケットはセッション中継部12-2の送信セッション処理部で処理される。

【実施例1】

【0038】

次に、本発明の第1の実施例によるセッション中継装置について説明する。本発明の第1の実施例によるセッション中継装置の構成は上記の図1に示す本発明の実施の形態によるセッション中継装置1と同様であり、その動作も本発明の実施の形態によるセッション中継装置1と同様である。

【0039】

図2は本発明の第1の実施例によるセッション中継装置の構成を示すブロック図である。図2において、セッション中継装置1はセッション識別部11と、セッション中継部12-1～12-Nと、パケットスケジューラ13と、出力制御部14とから構成されている。

【0040】

セッション識別部11は到着したパケットが属するセッションを決定する。セッション

中継部 12-1 ~ 12-N は送信端末 3 とのセッションと受信端末 2 とのセッションとの間で中継を行う。パケットスケジューラ 13 は各セッション中継部 12-1 ~ 12-N からのパケット出力を制御する。出力制御部 14 はパケットスケジューラ 13 からの指示に基づいて各セッション中継部 12-1 ~ 12-N からのパケット出力を行う。

#### 【0041】

また、セッション中継部 12-1 は受信端末 2 へとデータを送信するセッションの処理を行う送信セッション処理部 121-1 と、受信したデータを送信終了まで蓄えておく送信バッファ 122-1 と、送信端末 3 からデータを受信するセッションの処理を行う受信セッション処理部 123-1 とから構成されている。尚、図示していないが、セッション中継部 12-2 ~ 12-N の構成は上記のセッション中継部 12-1 の構成と同様である。

#### 【0042】

図 3 は図 2 のパケットスケジューラ 13 の構成を示すブロック図である。図 3 において、パケットスケジューラ 13 はリスト振り分け部 131 と、状態更新部 132 と、状態変数保存部 133 と、リセット制御部 134 と、送信可能リスト 135 と、送信待ちリスト 136 とから構成されている。

#### 【0043】

送信可能リスト 135 はパケット出力が可能なセッションの識別子を保持し、送信待ちリスト 136 は送信待ち状態にあるセッションの識別子を保持する。リスト振り分け部 131 はデータパケットセッションの識別子もしくは ACK パケットを受信したセッションの識別子を送信可能リスト 135 もしくは送信待ちリスト 136 に振り分ける。

#### 【0044】

状態更新部 132 は送信可能リスト 135 からセッションの識別子を取り出して出力制御部 14 に通知し、かつ該セッションの状態を更新する。状態変数保存部 133 は各セッションの状態を保持し、リセット制御部 134 は送信待ちリスト 136 で管理されているセッションの識別子を送信可能リスト 135 へと移動させる。

#### 【0045】

TCP (Transmission Control Protocol) セッションでは、通常、送信端末 3 と受信端末 2 との間の双方向の通信を行う。そのため、本実施例においては、一組の送信端末 3 及び受信端末 2 に対して 2 つのセッション中継部を使用するものとし、夫々の方向へのデータ通信に対しては夫々対応するセッション中継部を使用する。

#### 【0046】

したがって、セッション中継部 12-1 ~ 12-N は複数の送信端末 3 及び受信端末 2 の組に対して夫々 2 つずつ用意され、セッション中継部 12-1 ~ 12-N 各々は夫々の送信端末 3 からのセッションから対応する受信端末 2 へのセッションへと、もしくは夫々の受信端末 2 からのセッションから対応する送信端末 3 へのセッションへとデータの中継する処理を行う。

#### 【0047】

尚、TCP セッションにおいては、ある方向のデータパケットとその反対方向への ACK パケットとが 1 つのパケット上に統合される場合があるが (ACK の piggy back)、本実施例では説明を簡単化するために、このような動作に関しての説明については省略する。

#### 【0048】

図 4 及び図 5 は本発明の第 1 の実施例によるセッション中継装置 1 の動作を示すフローチャートであり、図 6 は図 3 に示すパケットスケジューラ 13 の動作を示すフローチャートである。これら図 1 ~ 図 6 を参照して本発明の第 1 の実施例によるセッション中継装置 1 の動作について説明する。

#### 【0049】

本実施例によるセッション中継装置 1 にパケットが入力された時 (図 4 ステップ S1)

、セッション識別部11はパケットのヘッダを参照し、送信元IP (Internet Protocol) アドレス、送信先IPアドレス、第四層プロトコル番号、送信元第四層ポート番号、送信先第四層ポート番号等に基づいて、パケットが属するセッションを決定する (図4ステップS2)。

#### 【0050】

セッション識別部11はパケットがデータパケットであれば (図4ステップS3)、対応するセッション中継部12-1~12-Nの受信セッション処理部123-1~123-N (受信セッション処理部123-2~123-Nは図示せず) へと渡す (図4ステップS4)。

#### 【0051】

また、セッション識別部11はパケットがACKパケットであれば (図4ステップS3)、対応するセッション中継部12-1~12-Nの送信セッション処理部121-1~121-N (送信セッション処理部121-2~121-Nは図示せず) へと渡す (図4ステップS10)。

#### 【0052】

受信セッション処理部123-1~123-Nでは入力されたデータパケットをシーケンス番号順に並べ替えて送信バッファ122-1~122-N (送信バッファ122-2~122-Nは図示せず) へと格納する (図4ステップS5)。また、受信セッション処理部123-1~123-Nは受信したデータパケットが正しいシーケンス番号を持つものである (図4ステップS6)、すなわち連続して受信しているデータの最後尾のシーケンス番号と連続していれば、送信端末3に対してデータの受信確認及び広告ウィンドサイズを通知するために、ACKパケットを出力制御部14を通して返送する (図4ステップS7)。

#### 【0053】

さらに、受信セッション処理部123-1~123-Nは連続して受信した最後のパケットのシーケンス番号を基にACKパケット (重複ACKパケット) を返送し、送信端末3に対してパケットの未到着を通知する。

#### 【0054】

上記の処理に関しては、TCP/IP Illustrated, Volume 1: The Protocols, Addison-Wesley, 1994, ISBN 0-201-63346-9 (以下、文献1とする) に詳しく記載されているので、その説明については詳述しない。

#### 【0055】

ACKパケットは生成された後、すぐさま出力制御部14から出力されるか、もしくは対応する逆方向のセッションの送信バッファ122-1~122-Nに格納され、逆方向のセッションのデータパケットとともに、パケットスケジューラ13の指示で出力される。

#### 【0056】

また、受信セッション処理部123-1~123-Nでは、連続して受信しているデータの最後尾のシーケンス番号をパケットスケジューラ13へと通知する (図4ステップS8)。

#### 【0057】

送信セッション処理部121-1~121-Nでは入力されたACKパケットを基に輻輳ウィンドサイズの変更を行い (図4ステップS9)、ACKパケットが重複ACKでなければ受信が確認されたデータを送信バッファ122-1~122-Nから消去し (図5ステップS12, S13)、重複ACKであれば必要に応じてデータの再送処理を行う (図5ステップS12, S11)。尚、この処理に関しても文献1に詳しく記載されているので、その説明については詳述しない。

#### 【0058】

送信セッション処理部121-1~121-Nは受信したACKパケットに記されてい

る広告ウインドと、更新した輻輳ウインドとをパケットスケジューラ13に通知する（図5ステップS15）。また、送信セッション処理部121-1～121-Nは再送がタイムアウトした場合（図5ステップS14）、受信確認済みの最後のシーケンス番号も通知する（図5ステップS16）。

#### 【0059】

尚、本実施例ではパケットスケジューラ13からパケット出力の指示があった時にのみパケットの出力を行うため、ここでは重複ACKを受信した時にすぐさまパケットの再送を行うのではなく、次に出力すべきパケットとして再送するパケットのシーケンス番号を記憶しておくのみである。

#### 【0060】

送信バッファ122-1～122-Nからのパケット出力は、パケットスケジューラ13からの指示に基づいて行われる。パケットスケジューラ13からパケット出力の指示があると、出力制御部14は送信バッファ122-1～122-Nからパケットを1つ取出して出力回線に出力し、出力したパケットのパケット長をパケットスケジューラ13に通知する。

#### 【0061】

送信バッファ122-1～122-Nから出力するパケットは、再送処理を行う場合に記憶しておいたシーケンス番号のパケットであり、さもなければ未送信のパケットのうち最もシーケンス番号の小さいパケットである。

#### 【0062】

尚、該セッションがTCPセッションでなければ、受信セッション処理部123-1～123-Nでは受信したパケットをそのまま到着順に送信バッファ122-1～122-Nに格納するのみである。送信セッション処理部121-1～121-NではACKパケットを受信せず、パケットスケジューラ13に対しては送信バッファ122-1～122-Nのキュー長を広告ウインド及び輻輳ウインドとして通知することで、送信バッファ122-1～122-Nにパケットがある限り、パケットスケジューラ13にパケット出力を要求し続ける。

#### 【0063】

パケットスケジューラ13では、状態変数保存部133において、セッション毎に、割り当てウェイト、送信可能バイト数、送信待ちバイト数の3つのパラメータを保持する。

#### 【0064】

割り当てウェイトは該セッションに割り当てられたウェイトである。パケットスケジューラ13は一定周期、あるいはパケット送信可能なセッションがなくなる毎にリセットを行い（図6ステップS26、S27）、このリセット1周期内に送信可能なバイト数が割り当てウェイトである。

#### 【0065】

送信可能バイト数は現時点から次のリセットまでに送信可能なバイト数であり、初期値は割り当てウェイトであって（図6ステップS21）、パケットを出力する毎にパケット長分だけ減算される（図6ステップS22、S23）。

#### 【0066】

送信待ちバイト数はTCPセッションの場合、TCPレイヤが送信許可したシーケンス番号から既にパケットスケジューラ13が送信を行ったシーケンス番号を引いたものであり、min（連続して受信しているデータの最後尾のシーケンス番号+1、受信端末が示した広告ウインド、該セッションの輻輳ウインド）-送信済みシーケンス番号となる。TCPセッションでない場合、送信待ちバイト数は送信バッファ122-1～122-N内のキュー長である。

#### 【0067】

図7は送信待ちバイト数を示す模式図である。図7において、パケット出力が可能なセッション、すなわち送信可能バイト数と送信待ちバイト数とがともに1あるいはMSS（Maximum Segment Size）以上のセッションの識別子は送信可能リス

ト135で管理され(図6ステップS24, S25)、送信可能バイト数がリセットされた後にパケット出力が可能となるセッション、すなわち送信待ちバイト数が1あるいはMSS以上あるが、送信可能バイト数が1あるいはMSS未満であるセッションの識別子は送信待ちリスト136で管理される(図6ステップS27~S29)。

**【0068】**

次に、図3を参照してパケットスケジューラ13の動作について説明する。リスト振り分け部131はセッション中継部12-1~12-Nから、送信端末3から連続して受信している最後尾のシーケンス番号や、受信端末2から受信した広告ウインド、更新した輻輳ウインドを受取ると、図7に示すようにして送信待ちバイト数を更新する。

**【0069】**

また、リスト振り分け部131は出力制御部14において再送タイマがタイムアウトした際にも、送信済みシーケンス番号を受信確認済みのシーケンス番号まで戻し、送信待ちバイト数を更新する。リスト振り分け部131は更新前の送信待ちバイト数が1あるいはMSS未満であれば、該セッションの識別子はまだ送信可能リスト135あるいは送信待ちリスト136で管理されていないため、更新後の送信待ちバイト数及び送信可能バイト数を基に該セッションの識別子を送信可能リスト135あるいは送信待ちリスト136に新たに格納する。

**【0070】**

状態更新部132は送信可能リスト135の先頭から送信可能なセッションの識別子を1つ取出し、これを出力制御部14に通知してパケット出力を行わせる。その後、状態更新部132は送信済みシーケンス番号に出力したパケット長を加算した後送信待ちバイト数を更新し、送信可能バイト数から出力したパケット長を減算する。

**【0071】**

状態更新部132は送信待ちバイト数及び送信可能バイト数にしたがって、改めて該セッションの識別子を送信可能リスト135あるいは送信待ちリスト136に格納する。すなわち、状態更新部132は送信待ちバイト数及び送信可能バイト数がともに1あるいはMSS以上であれば送信可能リスト135に格納し、送信待ちバイト数が1あるいはMSS以上であるが送信可能バイト数が1あるいはMSS未満であれば送信待ちリスト136に格納する。

**【0072】**

状態更新部132は送信可能リスト135が空になれば、全てのセッションの送信可能バイト数をリセット、すなわち送信可能バイト数に割り当てウエイトを加え、もし送信可能バイト数が割り当てウエイト以上となれば送信可能バイト数を割り当てウエイトの値とする[送信可能バイト数 = min (送信可能バイト数 + 割り当てウエイト, 割り当てウエイト)]。

**【0073】**

この処理によって、送信待ち状態にあったセッションは全て送信可能な状態に変化するため、送信待ちリスト136で管理されていたセッションを全て送信可能リスト135へと移動する。

**【0074】**

このリセット処理を行う前に、もし送信可能バイト数が1あるいはMSS未満ではなく、かつ該セッションの送信バッファ122-1~122-N内に送信待ちのデータがあれば、該セッションからの出力帯域に余裕はあるが、送信セッション処理部121-1~121-NにおいてTCP制御によって送信を止められている状態であるとする。

**【0075】**

この場合には、将来の送信再開に備えて帯域を蓄えておくため、送信可能バイト数と割り当てウエイトとの和が割り当てウエイトを超えた場合でも、ある一定値までは送信可能バイト数を蓄えておく[送信可能バイト数 = min (送信可能バイト数 + 割り当てウエイト, 送信可能バイト数上限値)]。

**【0076】**



上述したように、本実施例では、中継するパケットが受信時の送信バッファ122-1～122-Nへの格納と、送信時の送信バッファ122-1～122-Nからの取出しの際にのみ移動するため、データ移動のオーバーヘッドが小さい。また、TCPレイヤからの情報を用いたパケットスケジューラ13によってパケット出力制御を行うため、送信セッション処理部121-1～121-Nが直接パケット出力するよりも、少ない処理負荷で帯域制御を行いながらのパケット出力を行うことができる。

#### 【実施例2】

##### 【0077】

本発明の第2の実施例によるセッション中継装置の構成は図2に示す本発明の第1の実施例によるセッション中継装置1の構成と同様である。

##### 【0078】

図8は本発明の第2の実施例によるパケットスケジューラの構成を示すブロック図である。図8において、本発明の第2の実施例によるパケットスケジューラ16は割り当てウエイト変更部161及び制御パラメータ変更部162を加えた以外は図3に示す本発明の第1の実施例と同様の構成となっており、同一構成要素には同一符号を付してある。また、同一構成要素の動作は本発明の第1の実施例と同様である。

##### 【0079】

この図8を参照して本発明の第2の実施例によるパケットスケジューラ16の動作について説明する。ここでは、本発明の第1の実施例によるパケットスケジューラ13との相違点のみを説明する。

##### 【0080】

本実施例によるセッション中継装置では、パケットスケジューラ16内の制御パラメータ変更部162が割り当てウエイトとして設定した帯域にしたがって該セッションのTCP制御パラメータの値を動的に変更している。すなわち、制御パラメータ変更部162は該セッションの送信可能バイト数が割り当てウエイトよりも大きい予め設定した値よりも大きくなれば、TCP制御によってデータ出力が制限されていると判断し、該セッションがより多くの帯域でデータ出力が可能となるように、該セッションのTCP制御パラメータの値を変更する。

##### 【0081】

但し、TCP制御パラメータを変更しても該セッションのデータ出力帯域が増加しない場合、あるいは再送タイムアウトが一定頻度以上で発生する場合には、ネットワークの輻輳を防止するため、TCPパラメータの変更を停止する。

##### 【0082】

また、該セッションの送信可能バイト数が予め設定した別の値よりも小さくなれば、TCP制御によるデータ出力をパケットスケジューラ16の割り当てウエイトによる帯域まで減少させるため、該セッションのTCP制御パラメータの値を変更する。

##### 【0083】

前者の場合には、非輻輳時の輻輳ウインドを上げ幅を大きくしたり、輻輳時の輻輳ウインド下げ率を小さくしたりするが、後者の場合にはこの前者の場合とは逆の処理を行う。

##### 【0084】

また、本実施例によるセッション中継装置では、パケットスケジューラ16内の割り当てウエイト変更部161が現在送信可能な帯域にしたがって動的に割り当てウエイトを変更させている。すなわち、割り当てウエイト変更部161では該セッションの送信可能バイト数が割り当てウエイトよりも大きい予め設定した値よりも大きくなれば、割り当てウエイトで設定した帯域でのデータ出力が不可能であると判断し、割り当てウエイトを一時的に減少させる。

##### 【0085】

その後、割り当てウエイト変更部161は送信可能バイト数が予め設定した別の値よりも小さくなれば、割り当てウエイトを元の値まで順次増加させる。また、割り当てウエイト変更部161はセッション中継部12-1～12-Nから通知される該セッションの輻

輾ウインド及び往復伝播遅延時間計測値を基に該セッションが送信可能な帯域を計算する。

【0086】

さらに、割り当てウエイト変更部161はこの計算値と割り当てウエイトによる帯域設定値がある一定閾値以上異なれば、割り当てウエイトによる帯域設定値が上記の計算値となるように割り当てウエイトを一時的に変更する。

【実施例3】

【0087】

図9は本発明の第3の実施例によるセッション中継装置の構成を示すブロック図である。図9において、本発明の第3の実施例によるセッション中継装置は、セッション中継部15-1~15-Nに受信レート制御部151-1~151-N（受信レート制御部151-2~151-Nは図示せず）を設けた以外は図2に示す本発明の第1の実施例によるセッション中継装置と同様の構成となっており、同一構成要素には同一符号を付してある。また、同一構成要素の動作は本発明の第1の実施例と同様である。尚、受信レート制御部151-1~151-Nは送信端末3からの受信レートを制御する。

【0088】

この図9を参照して本発明の第3の実施例によるセッション中継装置の動作について説明するが、ここでは、上述した本発明の第1の実施例との相違点のみを説明する。

【0089】

本実施例によるセッション中継装置では、送信バッファ122-1~122-Nの空き容量がなくなると、受信セッション処理部123-1~123-Nが送信端末3に対して広告ウインドサイズが0であることを通知し、これに応じて送信端末3はデータの送信を停止する。

【0090】

本発明の第1の実施例では、その後、送信バッファ122-1~122-Nに空き容量ができたとしても、送信端末3がウインド検査のためのパケットを出力してくるまでは、送信再開のためのACKを出力することができない。

【0091】

これに対して、本実施例では、パケットスケジューラ13がパケット出力の指示を行った際、受信レート制御部151-1ではパケット出力後の送信バッファ122-1~122-Nの空き容量を検査し、もしこれが1あるいはMSS以上であれば、送信端末3に対してすばやい送信再開を促すためにACKパケットを生成する。

【0092】

また、受信レート制御部151-1~151-Nでは送信バッファ122-1~122-Nの空き容量、あるいはその平均値が予め設定したある値以上になれば、送信バッファ122-1~122-Nの空き容量がなくなること防ぐために、送信端末3に対して送信帯域の低下を指示する。

【0093】

これは、例えば、送信側に返送するACKパケットを重複ACKとすることや、受信パケットを廃棄すること、ACKパケットにECN (Explicit Congestion Notification) ビットを設定すること、ACKパケットを遅延させること、ACKパケットの広告ウインドを一時的に小さく書換えること等で可能である。

【実施例4】

【0094】

図10は本発明の第4の実施例によるセッション中継装置の構成を示すブロック図である。図10において、本発明の第4の実施例によるセッション中継装置は、送信iSCSI (internet Small Computer System Interface) 制御部171-1~171-N（送信iSCSI制御部171-2~171-Nは図示せず）及び受信iSCSI制御部172-1~172-N（受信iSCSI制御部172-2~172-Nは図示せず）をセッション中継部17-1~17-Nに設けた以外

は上述した本発明の第3の実施例によるセッション中継装置と同様の構成となっており、同一構成要素には同一符号を付してある。また、同一構成要素の動作は本発明の第3の実施例によるセッション中継装置と同様である。

【0095】

送信 iSCSI 制御部 171-1 ~ 171-N は受信端末 2 への送信レートに対して iSCSI レイヤの輻輳制御情報を反映させる。受信 iSCSI 制御部 172-1 ~ 172-N は送信端末 3 からの受信レートに対して iSCSI レイヤの輻輳制御情報を反映させる。

【0096】

この図 10 を参照して本発明の第 4 の実施例によるセッション中継装置の動作について説明する。ここでは、本発明の第 4 の実施例における本発明の第 3 の実施例との相違点のみを説明する。

【0097】

iSCSI レイヤにおいては、受信端末 2 から送信端末 3 に対して、受信端末 2 が受信可能なデータ量を通知する R2T (Ready-To-Transfer) パケットを送信することで、送信端末 3 と受信端末 2 との間で転送制御を行う。

【0098】

そこで、本実施例では、送信 iSCSI 制御部 171-1 ~ 171-N が受信端末 2 から R2T パケットを受信すると、この R2T パケットを受信端末 2 に送らず、本実施例によるセッション中継装置にて R2T パケット受信時に出力を行ったデータの最後のシーケンス番号を記憶しておく。

【0099】

送信 iSCSI 制御部 171-1 ~ 171-N は前回記憶しておいたシーケンス番号に、今回受信した R2T パケットの送信可能データ量を加えたものを、iSCSI レイヤで送信可能なデータ量としてパケットスケジューラ 13 に通知する。

【0100】

パケットスケジューラ 13 では送信待ちバイト数を、TCP レイヤが送信許可したデータ量と iSCSI レイヤが許可したデータ量との最小値とする。すなわち、図 11 に示すように、送信待ちバイト数 = min (連続して受信しているデータの最後尾のシーケンス番号 + 1, 受信端末が示した広告ウインド, 該セッションの輻輳ウインド, 前回 R2T パケットを受信した際のシーケンス番号 + 今回受信した R2T パケットの送信可能データ量) - 送信済みシーケンス番号となる。

【0101】

受信 iSCSI 制御部 172-1 ~ 172-N ではパケットスケジューラ 13 がパケット出力の指示を行った際に、パケット出力後の送信バッファ 122-1 ~ 122-N の空き容量を検査し、もしこれが予め定められ一定値以上であれば、送信端末 3 に対して送信バッファの空き容量を R2T パケットとして送信する。

【0102】

上述したように、本実施例では送信端末 3 と受信端末 2 との間で iSCSI レイヤの輻輳制御を行うのではなく、セッション中継装置で iSCSI レイヤの輻輳制御を中継することで、iSCSI レイヤにおいても TCP セッションを中継するのと同様のスループットの向上を図ることができる。

【実施例 5】

【0103】

図 12 は本発明の第 5 の実施例によるセッション中継装置の構成を示すブロック図である。図 12 において、本発明の第 5 の実施例によるセッション中継装置はセッション識別部 11 と、セッション送信部 41-1 ~ 41-N と、セッション受信部 42-1 ~ 42-N と、パケットスケジューラ 13 と、出力制御部 14 とから構成されている。

【0104】

セッション識別部 11 は到着したパケットが属するセッションを決定し、セッション送

信部41-1~41-Nは受信端末へのセッションのデータ送信処理を行い、セッション受信部42-1~42-Nは受信端末へのセッションからのデータ受信処理を行う。

【0105】

パケットスケジューラ13は各セッション送信部41-1~41-Nからのパケット出力を制御する。出力制御部14はパケットスケジューラ13からの指示に基づいて各セッション送信部41-1~41-Nからのパケット出力を行う。

【0106】

また、セッション送信部41-1~41-Nは送信セッション処理部411-1~411-N（送信セッション処理部411-2~411-Nは図示せず）と、送信データ生成部412-1~412-N（送信データ生成部412-2~412-Nは図示せず）と、送信バッファ413-1~413-N（送信バッファ413-1~413-Nは図示せず）とから構成されている。

【0107】

送信データ生成部413-1~413-Nはアプリケーションプログラムからの送信データを送信バッファ413-1~413-Nへと格納する。送信バッファ5-6-1は送信するデータを一時蓄えておく。送信セッション処理部411-1~411-Nは受信端末へとデータを送信するセッションの処理を行う。

【0108】

セッション受信部42-1~42-Nは受信セッション処理部421-1~421-N（受信セッション処理部421-2~421-Nは図示せず）と、受信バッファ422-1~422-N（受信バッファ422-2~422-Nは図示せず）と、受信データ処理部423-1~423-Nとから構成されている。

【0109】

受信セッション処理部421-1~421-Nは受信端末からのデータの受信処理を行い、受信バッファ422-1~422-Nは受信したデータを一時蓄えておく。受信データ処理部423-1~423-Nは受信バッファ5-14-1からアプリケーションへ受信データを受け渡す。

【0110】

TCPセッションでは、通常、送信端末と受信端末との間の双方向の通信を行うため、本実施例では、一組の送信端末及び受信端末に対してそれぞれ1つのセッション送信部41-1~41-N及びセッション受信部42-1~42-Nを使用する。本実施例においては、セッション中継装置が送信端末あるいは受信端末を兼ねている。

【0111】

図13は本発明の第5の実施例によるセッション中継装置（送信端末）及びセッション中継装置（受信端末）の間のデータの流れを示すブロック図である。図13において、セッション中継装置（送信端末）4-2からセッション中継装置（受信端末）4-1へデータを送る場合、セッション中継装置（送信端末）4-2のセッション送信部41-1-2から出力されたデータパケットはセッション中継装置（受信端末）4-1のセッション受信部42-1-1で受信処理が行われる。その結果、生成されたACKパケットはセッション中継装置（送信端末）4-2のセッション送信部41-1-2へと返送される。

【0112】

また、セッション中継装置（受信端末）4-1からセッション中継装置（送信端末）4-2へデータを送る場合、セッション中継装置（受信端末）4-1のセッション送信部41-1-1から出力されたデータパケットはセッション中継装置（送信端末）4-2のセッション受信部42-1-2で受信処理が行われる。その結果、生成されたACKパケットはセッション中継装置（受信端末）4-1のセッション送信部41-1-1へと返送される。

【0113】

次に、図12を参照して本発明の第5の実施例の動作について説明する。まず、送信端末から受信端末へのデータ転送に関して説明する。

## 【0114】

アプリケーションプログラムが出力した送信データは送信データ生成部412-1~412-Nによって送信バッファ413-1~413-Nへと書込まれる。送信セッション処理部411-1~411-Nは送信バッファ413-1~413-Nから受信端末へのデータ送信処理を行う。この処理は上述した本発明の第1の実施例と同様であるため、その説明を省略する。

## 【0115】

ここで、パケットスケジューラ13は送信待ちバイト数を計算する際に、連続して受信しているデータの最後尾のシーケンス番号の代わりに、アプリケーションプログラムから受取ったデータの最後のシーケンス番号を用いる。

## 【0116】

次に、受信端末から送信端末へのデータ転送に関して説明する。受信セッション処理部421-1~421-Nでは受信端末から送信されたデータの受信処理を行い、正しく受取ることができたデータを受信バッファ422-1~422-Nに格納する。

## 【0117】

この受信処理は格納するバッファが送信バッファではなく受信バッファである点を除いて、上述した本発明の第1の実施例と同様であるため、その説明を省略する。受信バッファ422-1~422-Nに書込まれたデータは、受信データ処理部423-1~423-Nによって取出されてアプリケーションプログラムへと渡される。

## 【0118】

このように、本発明は、複数のレイヤにおける輻輳制御処理に関して、輻輳制御情報の作成のみをそれぞれのレイヤで行い、パケット出力制御処理をIPレイヤのスケジューラに集約することで、セッション中継処理の高速化を実現することができる。

## 【0119】

また、本発明は、複数のレイヤの受信バッファ及び送信バッファをIPレイヤの送信バッファに集約することによって、バッファ間でのデータ移動をなくし、セッション中継処理の高速化を実現することができる。

## 【図面の簡単な説明】

## 【0120】

【図1】本発明の実施の形態によるセッション中継装置の構成を示すブロック図である。

【図2】本発明の第1の実施例によるセッション中継装置の構成を示すブロック図である。

【図3】図2のパケットスケジューラの構成を示すブロック図である。

【図4】本発明の第1の実施例によるセッション中継装置の動作を示すフローチャートである。

【図5】本発明の第1の実施例によるセッション中継装置の動作を示すフローチャートである。

【図6】図3に示すパケットスケジューラの動作を示すフローチャートである。

【図7】送信待ちバイト数を示す模式図である。

【図8】本発明の第2の実施例によるパケットスケジューラの構成を示すブロック図である。

【図9】本発明の第3の実施例によるセッション中継装置の構成を示すブロック図である。

【図10】本発明の第4の実施例によるセッション中継装置の構成を示すブロック図である。

【図11】本発明の第4の実施例における送信待ちバイト数を説明する模式図である。

【図12】本発明の第5の実施例によるセッション中継装置の構成を示すブロック図である。

【図13】本発明の第5の実施例によるセッション中継装置（送信端末）及びセッション中継装置（受信端末）の間のデータの流れを示すブロック図である。

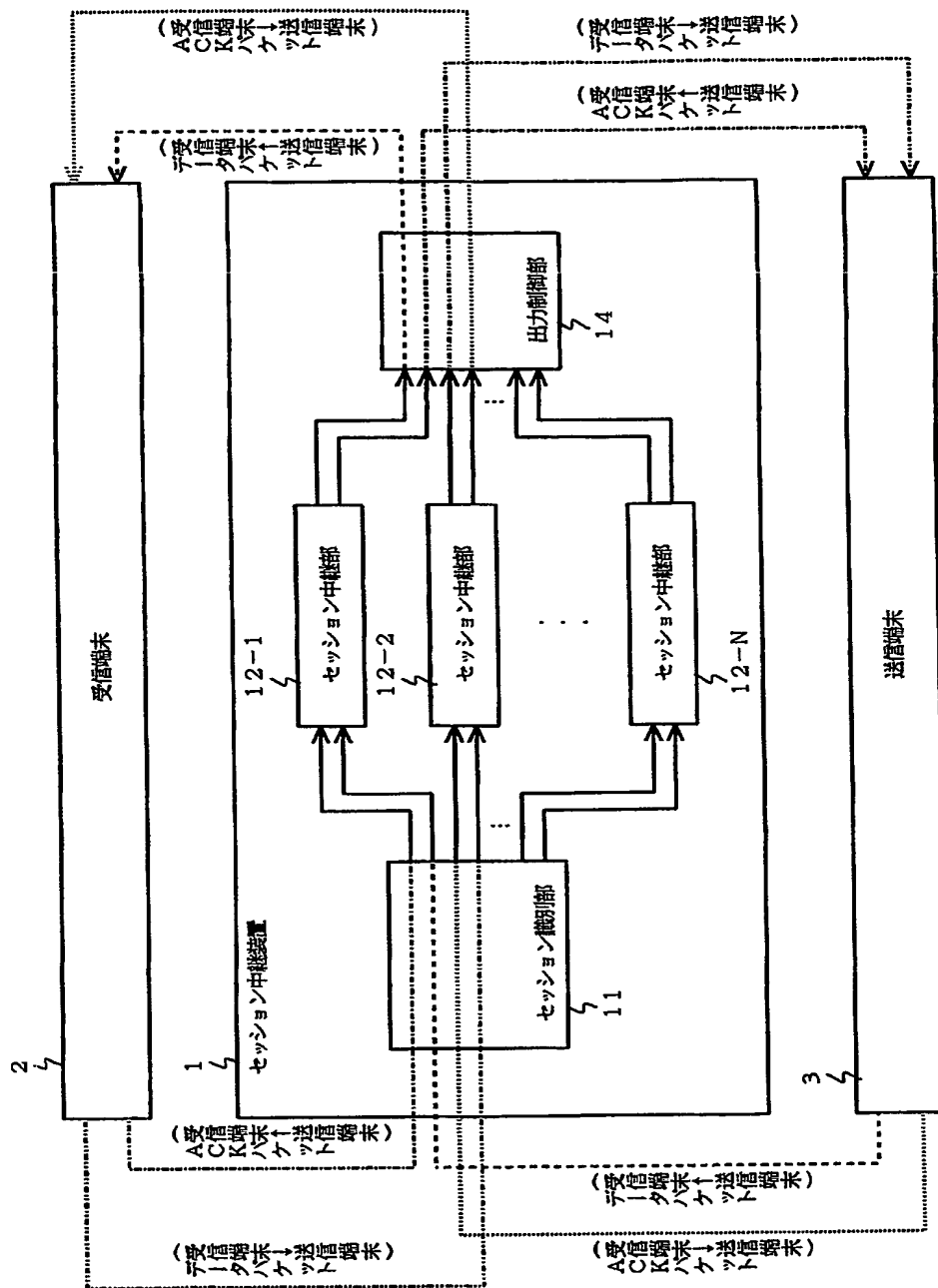
## 【符号の説明】

【0121】

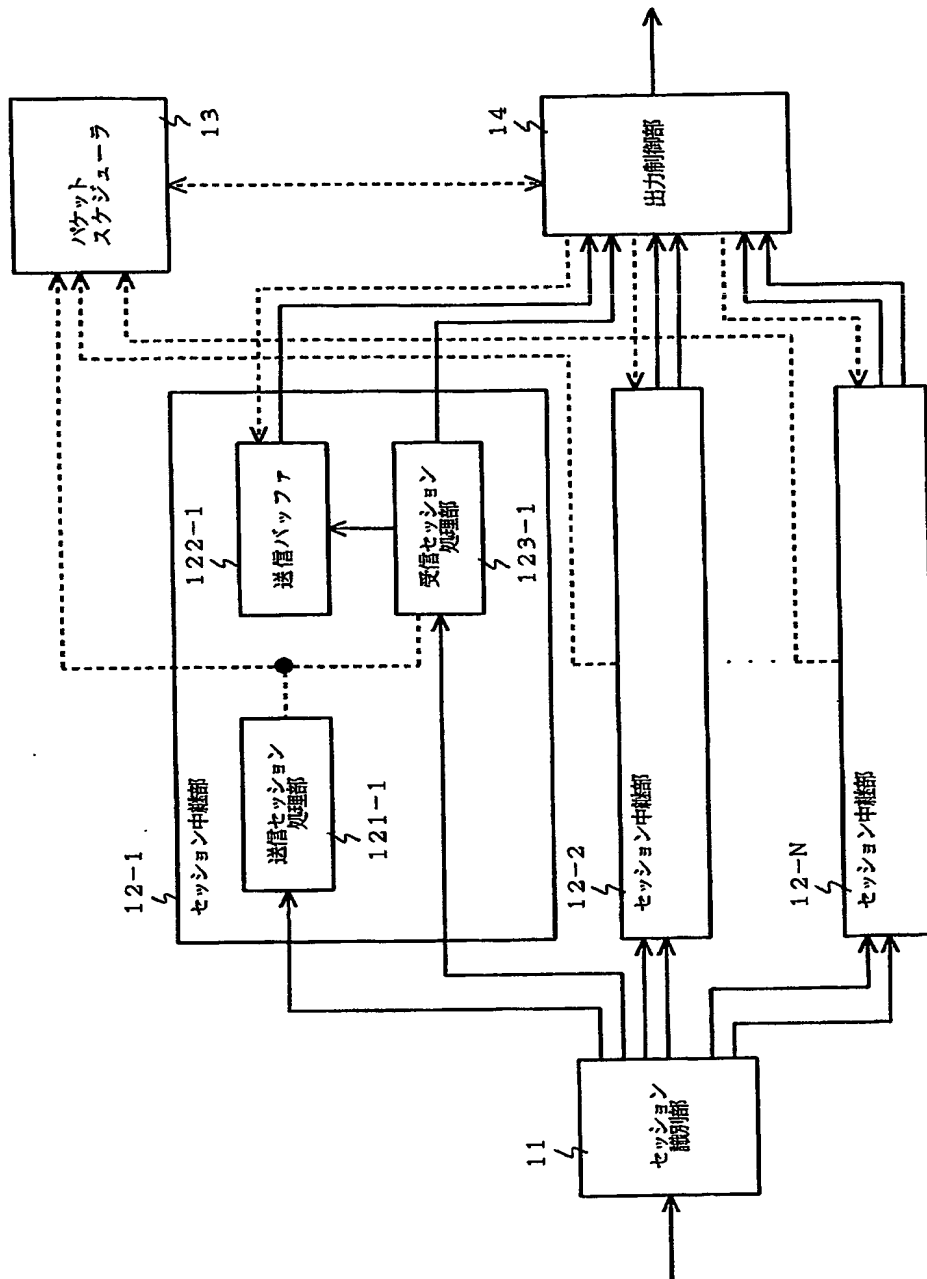
- 1 セッション中継装置
- 2 受信端末
- 3 送信端末
- 4-1 セッション中継装置（受信端末）
- 4-2 セッション中継装置（送信端末）
- 11, 11-1, 11-2 セッション識別部
- 12-1~12-N,
- 15-1~15-N,
- 17-1~17-N セッション中継部
- 13, 16 パケットスケジューラ
- 14, 14-1, 14-2 出力制御部
- 41-1~41-N,
- 41-1-1, 41-1-2 セッション送信部
- 42-1~42-N,
- 42-1-1, 42-1-2 セッション受信部
- 121-1, 411-1 送信セッション処理部
- 122-1, 413-1 送信バッファ
- 123-1, 421-1 受信セッション処理部
- 131 リスト振り分け部
- 132 状態更新部
- 133 状態変数保存部
- 134 リセット制御部
- 135 送信可能リスト
- 136 送信待ちリスト
- 151-1 受信レート制御部
- 161 割り当てウエイト変更部
- 162 制御パラメータ変更部
- 171-1 送信iSCSI制御部
- 172-1 受信iSCSI制御部
- 412-1 送信データ生成部
- 422-1 受信バッファ
- 423-1 受信データ処理部

【書類名】 図面

【図1】

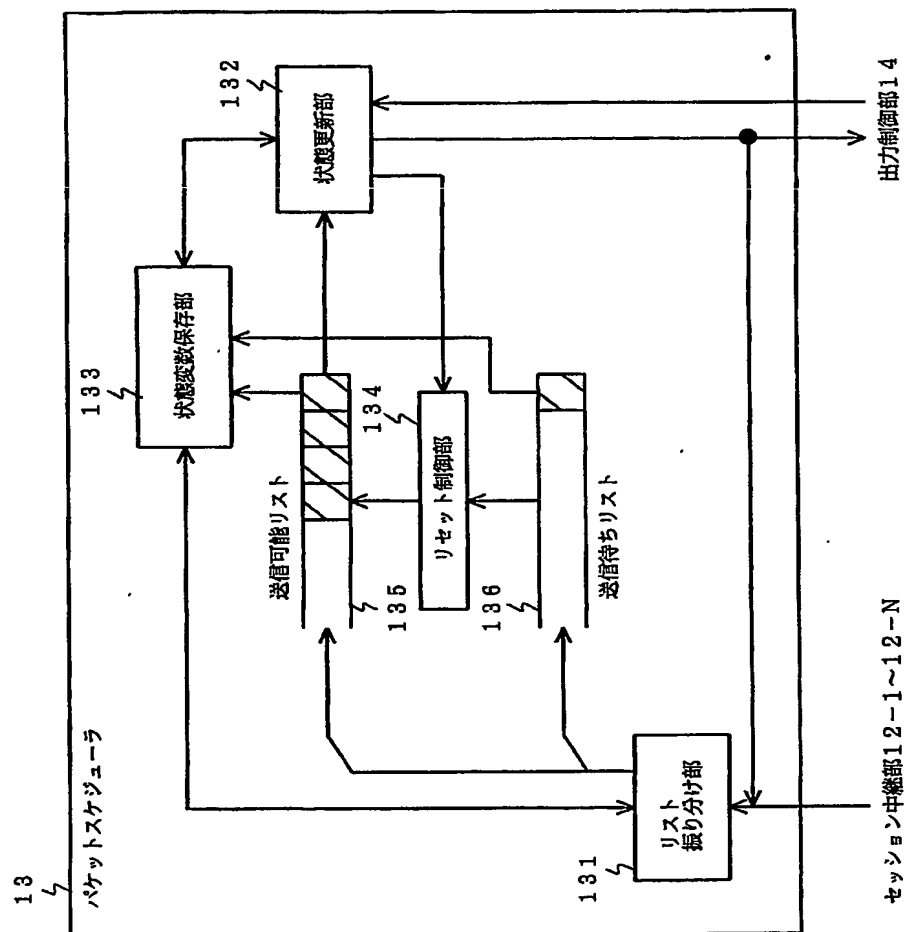


【図 2】

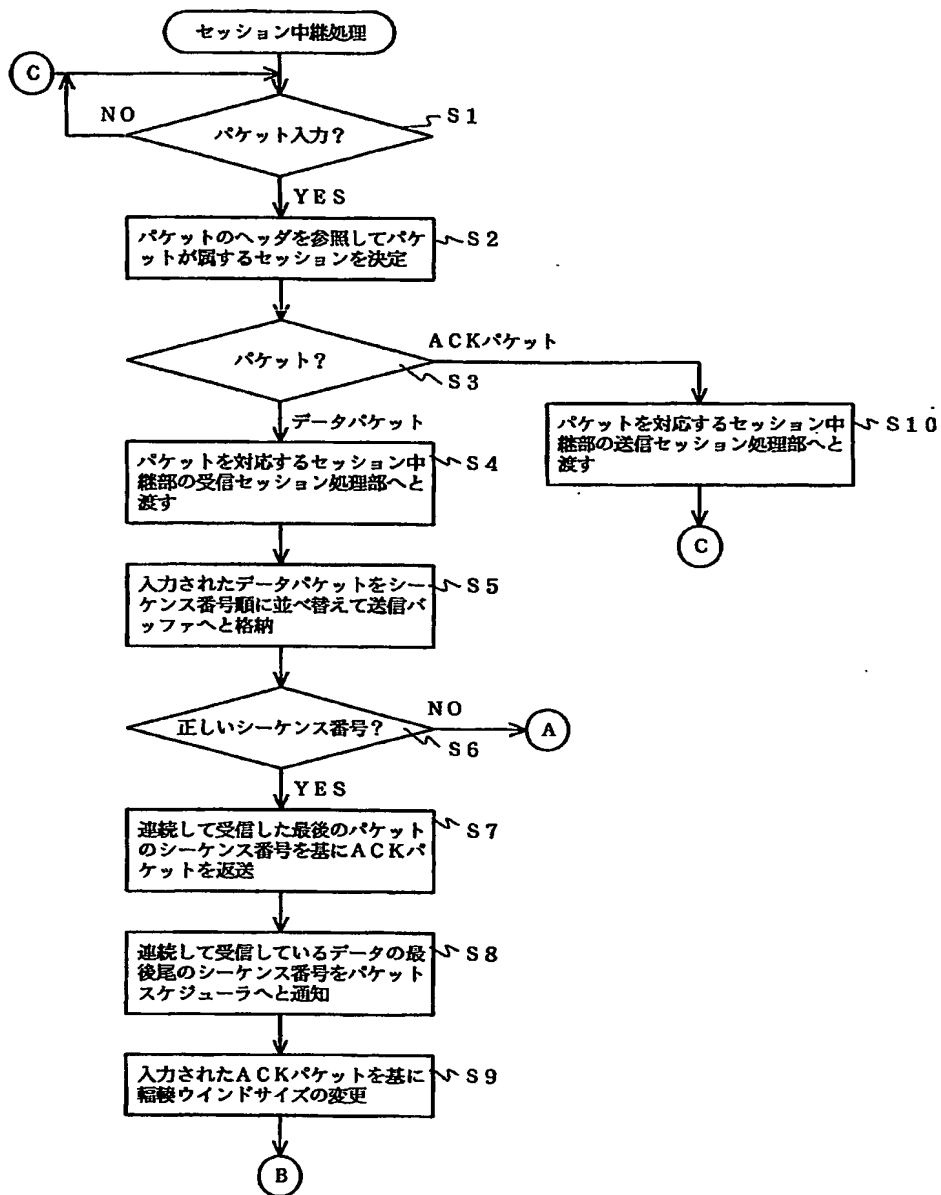




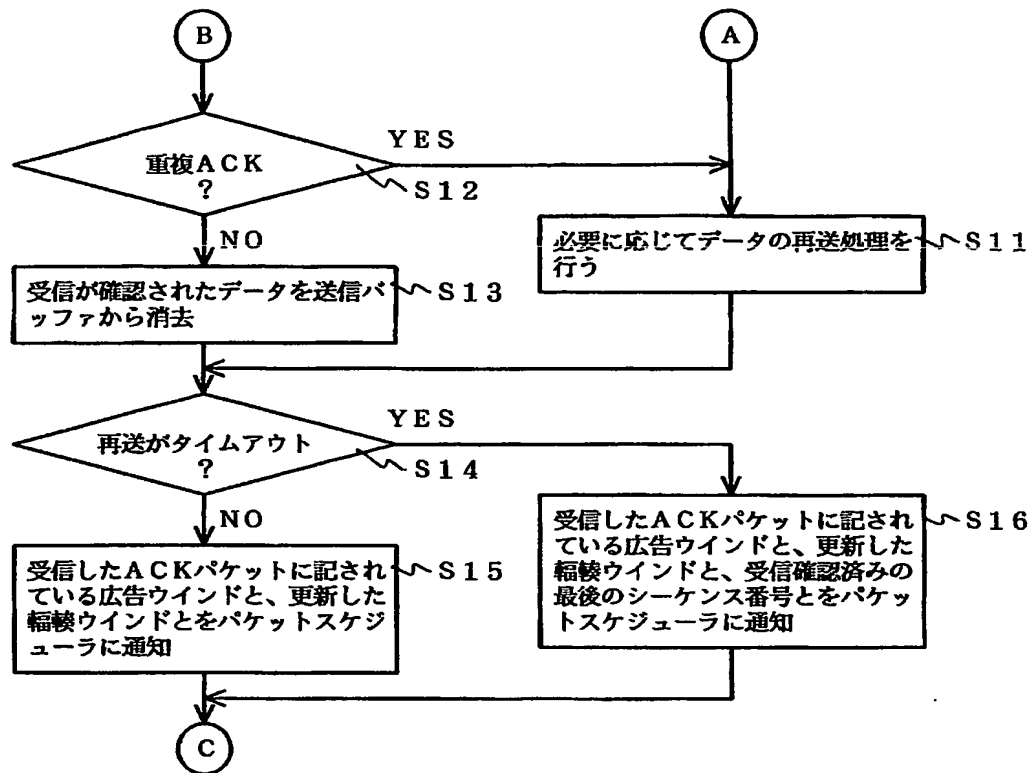
【図 3】



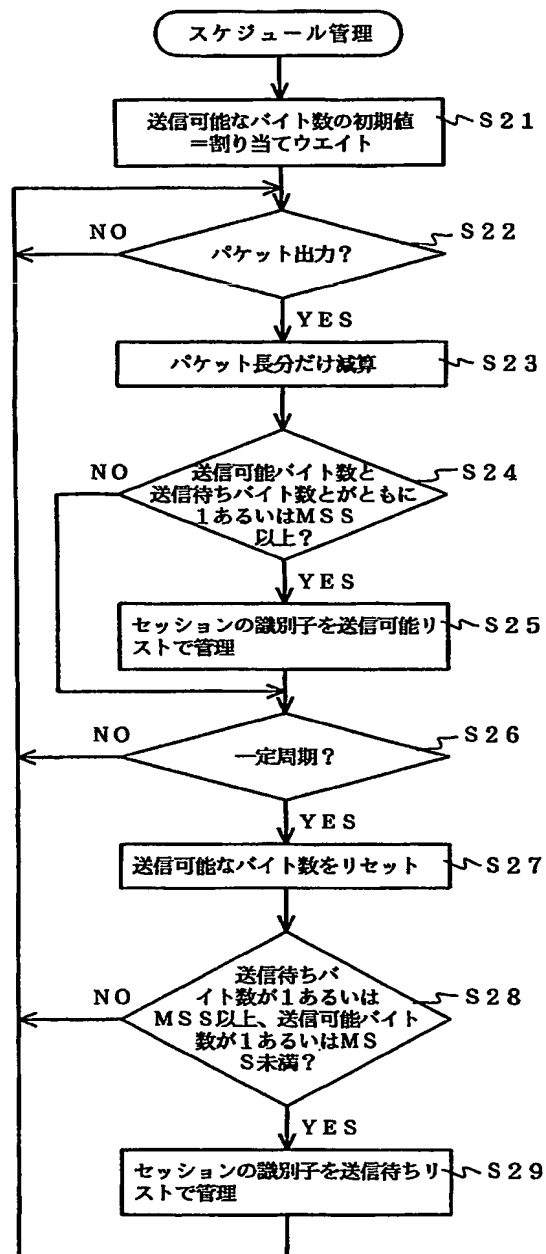
【図 4】



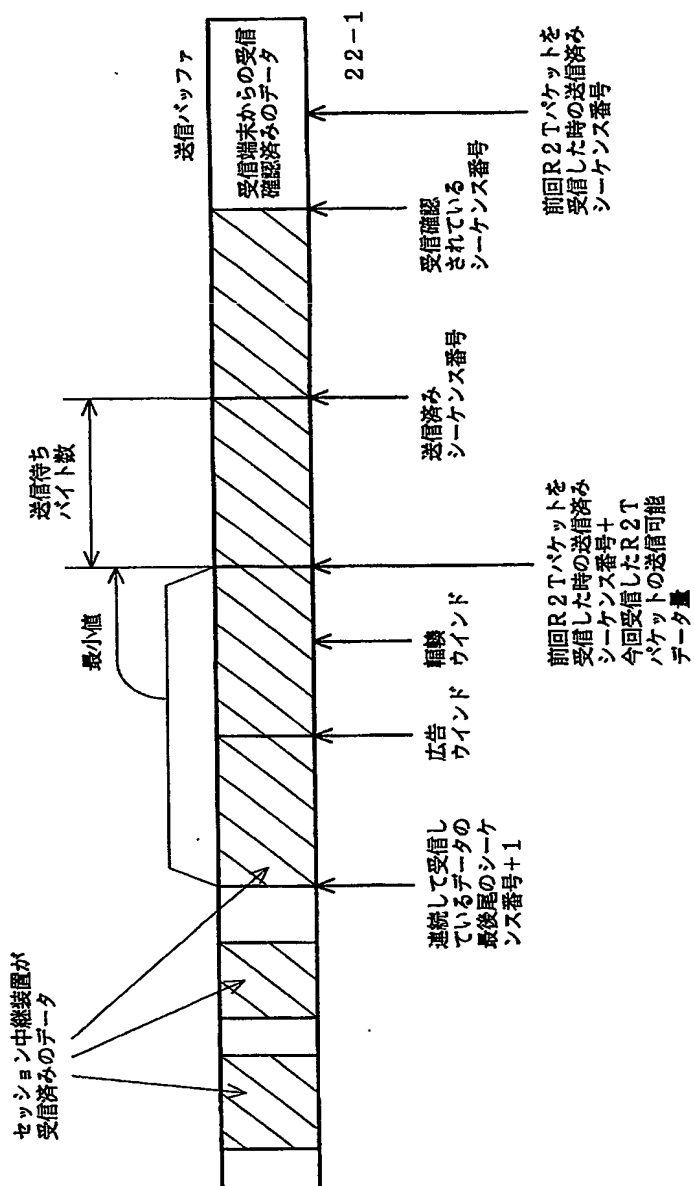
【図 5】



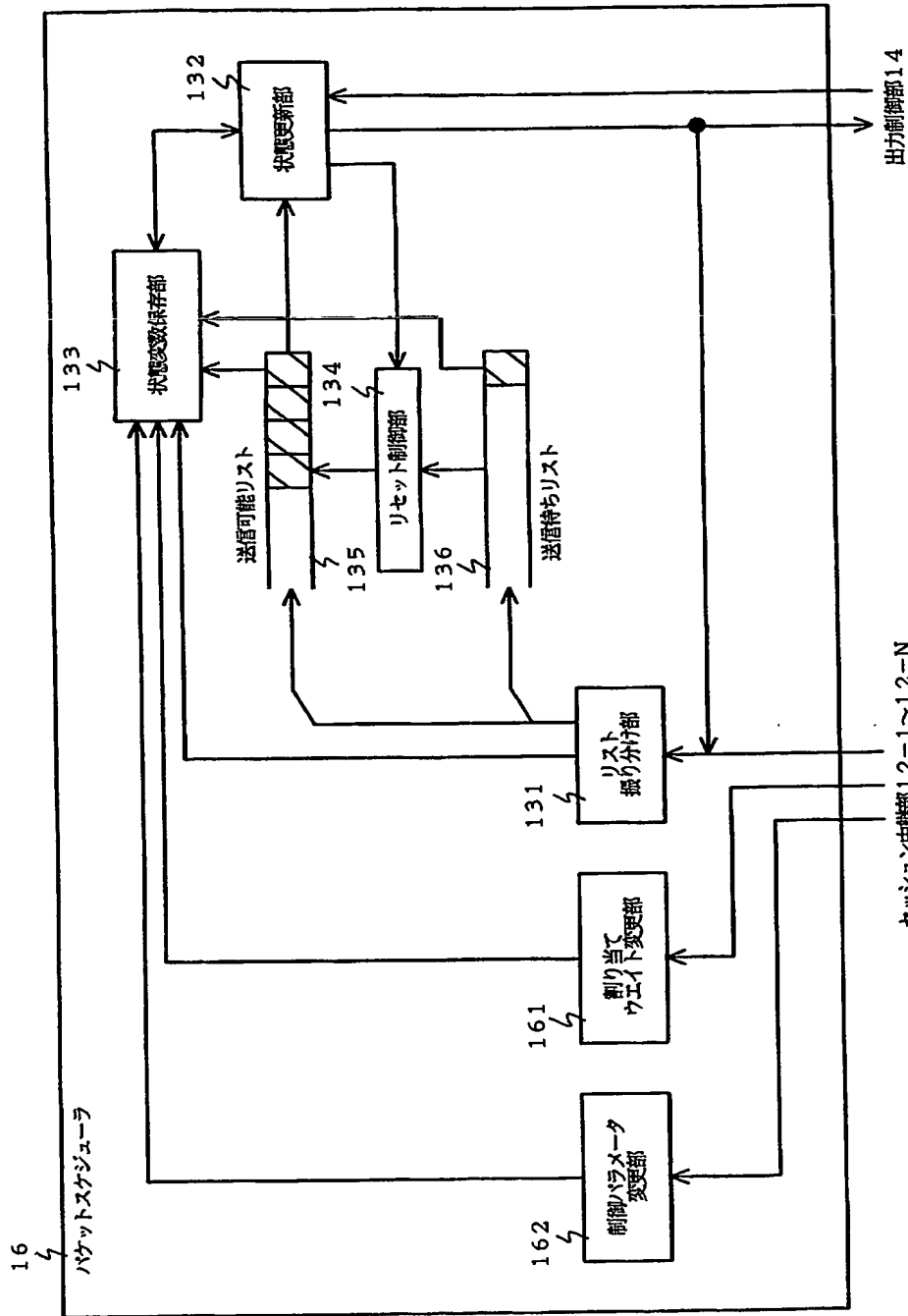
【図 6】



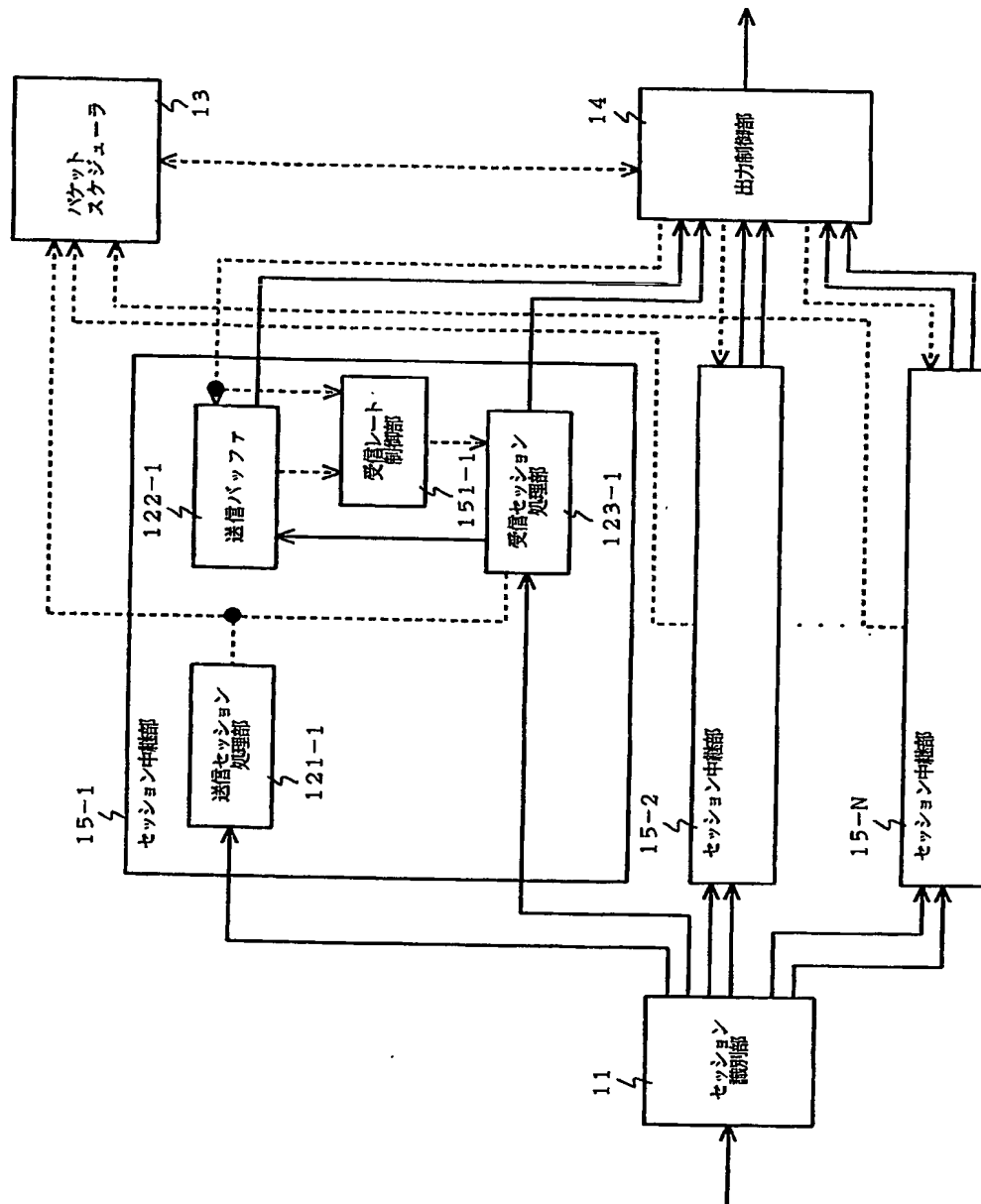
【図 7】



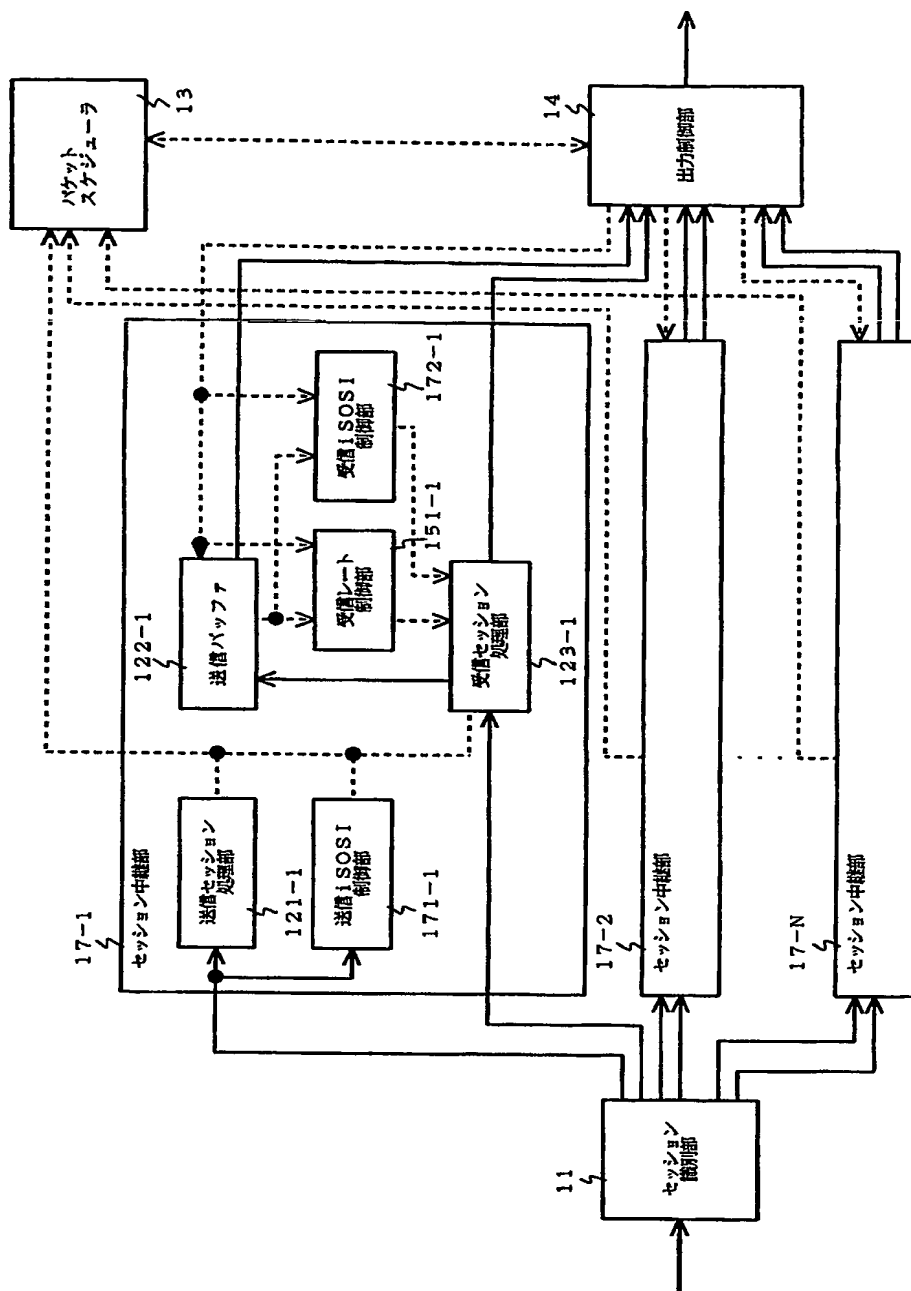
【図 8】



【図 9】

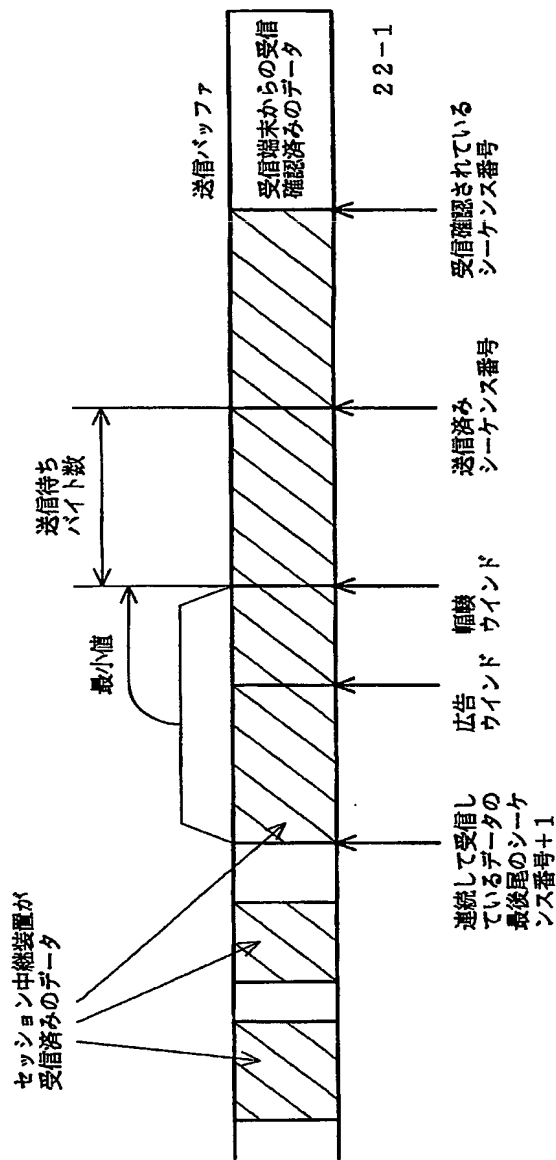


【図 10】

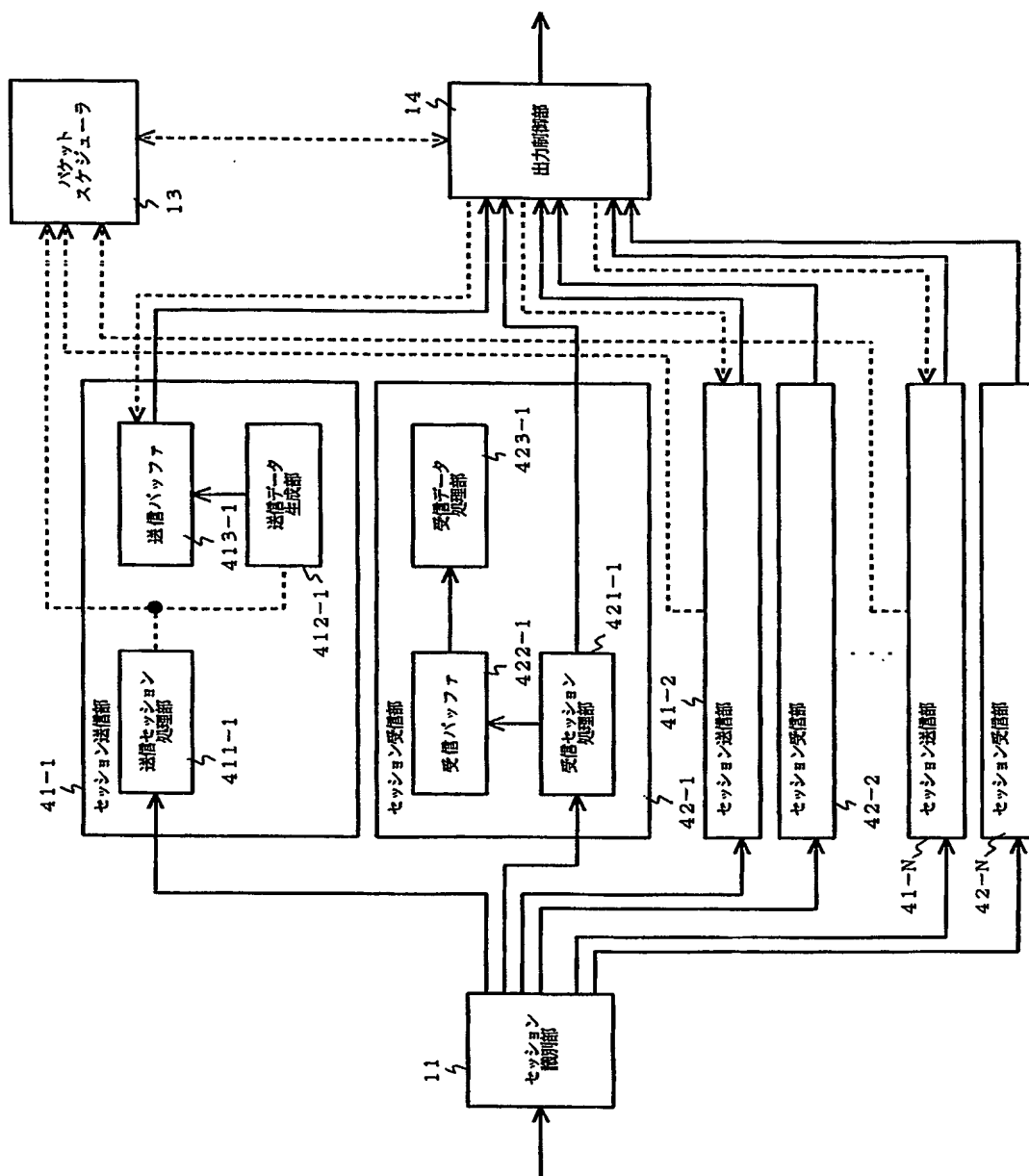




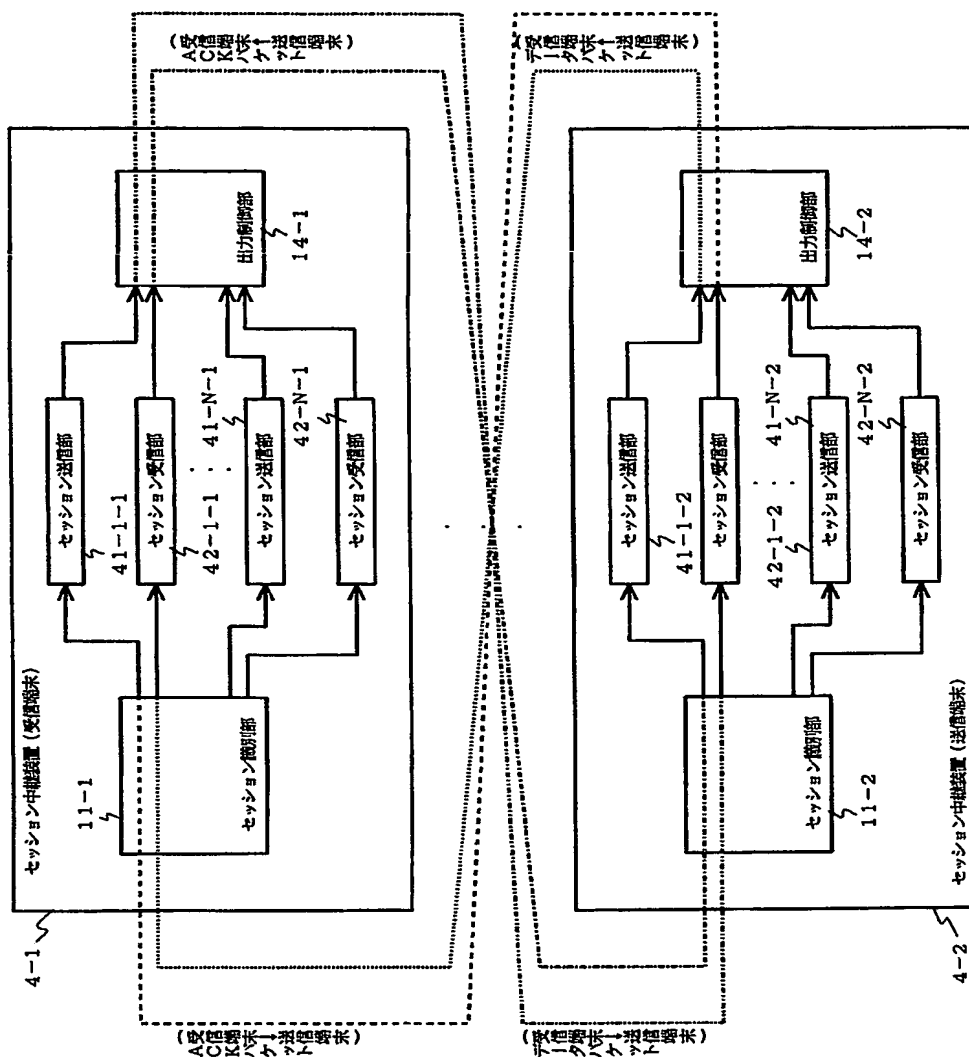
【図 11】



【図12】



【図 13】



**【書類名】 要約書****【要約】**

**【課題】** 中継するセッション数が多く、パケットの出力制御を行う場合でも、高速にセッション間の中継処理を行えるセッション中継装置を提供する。

**【解決手段】** セッション識別部 11 は到着したパケットが属するセッションを決定し、セッション中継部 12-1 ~ 12-N は送信端末とのセッションと受信端末とのセッションとの間で中継を行う。パケットスケジューラ 13 は各セッション中継部 12-1 ~ 12-N からのパケット出力を指示制御し、出力制御部 14 はその指示に基づいて各セッション中継部 12-1 ~ 12-N からのパケット出力を行う。セッション中継部 12-1 において、送信セッション処理部 121-1 は受信端末へとデータを送信するセッションの処理を行い、送信バッファ 122-1 は受信したデータを送信終了まで蓄えておく。受信セッション処理部 123-1 は送信端末からデータを受信するセッションの処理を行う。

**【選択図】** 図 2

特願 2 0 0 3 - 2 9 5 7 5 8

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 4 2 3 7 ]

1. 変更年月日

1 9 9 0 年 8 月 2 9 日

[変更理由]

新規登録

住 所

東京都港区芝五丁目 7 番 1 号

氏 名

日本電気株式会社